

538048-1

538048  
(15)  
2a- 1

To appear in the *Journal of Applied Meteorology*, 2000.

## A Regularized Neural Net Approach for Retrieval of Atmospheric and Surface Temperatures with the IASI Instrument

F. Aires<sup>1</sup>, A. Chédin, N.A. Scott

Laboratoire de Météorologie Dynamique, École Polytechnique, France

and W.B. Rossow

NASA Goddard Institute for Space Studies, New-York

### Abstract

In this paper, a fast atmospheric and surface temperature retrieval algorithm is developed for the high resolution Infrared Atmospheric Sounding Interferometer (IASI, space-borne instrument). This algorithm is constructed on the basis of a neural network technique that has been regularized by introduction of a priori information. The performance of the resulting fast and accurate inverse radiative transfer model is presented for a large diversified dataset of radiosonde atmospheres including rare events. Two configurations are considered: a tropical-airmass specialized scheme and an all-air-masses scheme.

## 1. Introduction

The Infrared Atmospheric Sounding Interferometer (IASI) is a high resolution ( $0.25\text{ cm}^{-1}$ ) Fourier transform spectrometer scheduled for flight in 2005 on the European polar METeorological Operational Platform (METEOP-1) satellite funded by the European organization for METeorological SATellites (EUMETSAT) and the European Space Agency (ESA) member states. This instrument is intended to replace the High Resolution Infrared Radiation Sounder (HIRS) as the operational infrared sounder and is expected to reach accuracies of 1 K in temperature and 10 % in water vapor with vertical resolutions of 1 km and 2 kms respectively. IASI, jointly developed by the Centre National d'Études Spatiales (CNES) and EUMETSAT, provides spectral coverage from  $3.5\ \mu\text{m}$  and  $15.5\ \mu\text{m}$  at considerably higher spectral resolution than HIRS and, together with the Advanced Microwave Sounding Unit (AMSU), is expected to lead to dramatic improvements in the accuracy and height resolution of remotely sensed temperature and humidity profiles and ozone amount.

The goal of this study is to present an inversion algorithm that retrieves geophysical variables from IASI measurements. We are confronted, in this work, with problems related to the ill-posed character of the inverse problem, the sensitivity to noise and, specific to IASI, the data dimension. The Multi-Layer Perceptron (MLP) technique is particularly interesting to solve this kind of problems. Such an approach has already been developed by the Atmospheric Radiation Analysis (ARA) group of LMD for HIRS coupled with the Microwave Sounding Unit (MSU) [Escobar *et al.*, 1993], for the Special Sensor Microwave / Temperature (SSM/T) instrument on board the Defense Meteorological Satellite Program (DMSP) [Rieu *et al.*, 1996], and even for the high resolution infrared spectrometer Advanced Infrared Radiation Sounder (AIRS) of National Aeronautics and Space Administration (NASA) for the coming Earth Observation System (EOS-PM-1) [Escobar *et al.*, 1993] or for IASI instrument [Aires *et al.*, 1998]. The great advantages of MLP are the rapidity, the small amount of memory required and accuracy of results [Aires, 1999]. The MLP model is nonlinear, which is a crucial point for the regression fit to the inverse Radiative Transfer Equation (RTE). Furthermore, assumptions like the linearity of the RTE or the Gaussian assumptions for stochastic variables, are not required for the MLP.

In this paper, it is demonstrated that the inver-

sion procedure can be regularized by introducing various kind of *a priori* information about the physical problem to the neural method. This may be done within the three components of the neural network technique: the architecture of the network, the learning algorithm and the learning data base. This approach overcomes the "black-box" modeling conception often associated with Neural Network methods.

We present here an application to the problem of surface temperature and the atmospheric temperature profile retrievals with the IASI instrument. Previous studies have used information content analysis to estimate the expected retrieval errors of IASI [Amato and Serio, 1997; Prunet *et al.*, 1998]; but this kind of estimate is dependent on some assumptions (Gaussian hypothesis, independence of first-guess and observation, first-guess error covariance matrices often taken to be diagonal, i.e. no correlations among the first-guess errors of the variables, etc), and in the limited number of atmospheric situations that have been examined.

Our neural network model is learned and tested through a large number, 3500, of real atmospheric situations as measured by radiosondes, taken from the Thermodynamic Initial Guess Retrieval (TIGR) data base [Chédin *et al.*, 1985; Achard, 1991; Escobar, 1993b; Chevallier *et al.*, 1998; 2000]. These atmospheric situations include very complex temperature profiles that are often much more irregular than re-analysis data or model output data. Rare situations are also included so that the dataset represents, as much as possible, all kinds of possible atmospheric situations (initially for a pattern recognition purpose). This complexity represents a higher variability than that encountered in operational conditions with model output data, so our estimation of the retrieval errors could be an over-estimate. However the use of a large and complex climatological dataset allows the inversion model to be calibrated globally and even for rare events. Furthermore, our analysis of the retrieval error is made for realistic instrumental noise conditions. Contrary to other approaches, no assumptions about the physical problem are used, like the linear or the Gaussian assumptions.

This paper is organized as follows. The physical problem associated with our application is presented in section 2. The neural network approach is described in section 3. The data bases used in this study are presented in section 4. Two applications of our neural technique are then presented: the surface temperature retrieval (section 5) and the atmospheric temper-

ature profile retrieval (section 6). Short conclusions and perspectives are given in section 7.

## 2. Sounding the Atmosphere with the IASI Instrument

### 2.1. Radiative Transfer in the Atmosphere

The radiance measured by an instrument at the top of the atmosphere depends on the atmospheric and surface properties. This dependence is described by the Radiative Transfer Equation (RTE):

$$I(\nu) = \varepsilon_s B(T_s, \nu) \tau_{s\nu} + \int_{P_s}^{P_{top}} B(T(P), \nu) \frac{\partial \tau_\nu}{\partial \ln P}(P) d \ln P \quad (1)$$

where  $\nu$  is the wavenumber ( $cm^{-1}$ ),  $\varepsilon_s$  the Earth's surface emissivity which may be a function of wavenumber,  $B(T(P), \nu)$  the Planck function which indicates the radiance emitted by a black-body at temperature  $T$  and atmospheric pressure  $P$ ,  $\tau_\nu$  the transmission factor between the satellite and the pressure level  $P$ .  $\frac{\partial \tau_\nu}{\partial \ln P}(\cdot)$  is often referred to as the weighting function. The RTE expresses the two radiative contributions at the top of the atmosphere: one arising from the surface (first term in right hand side) and one from the atmosphere (second term in right hand side). The equation's complexity lies in the transmission factors which depend on pressure, temperature, concentration of gases, spectroscopic characteristics of the absorbing gases ( $CO_2$ ,  $H_2O$ ,  $O_3$ , ...).

To retrieve atmospheric variables from radiative measurements at the top of the atmosphere, the inverse of equation (1) has to be computed. The analytical inversion of this function is not possible, only an inference approach can be used [Twomey, 1977]. Contrary to the direct problem, which can advantageously be estimated with high precision by a physical algorithm, the inverse problem needs a method of resolution based on a statistical representation of the (unknown) inverse equation. Two general approaches exist: use an inversion scheme for each observation (we call this approach the *local inversion*) or model the inverse RTE once and for all (we call this approach the *global inversion*). The local inversion requires generally a good initial guess solution and a rapid and accurate direct transfer model [Rodgers, 1976]. Even if global inversion models can use a first-guess [Aires et al., 2000], this is not required and no direct model is used during operational use. Although global inversion does not have these two limitations, it is a more ambitious problem.

### 2.2. Instrumental Characteristics

The two major advances of the IASI instrument are:

- The dramatically increased number of spectral channels: for each field of view, 8461 measures are available (covering the spectral range from 645 to 2760  $cm^{-1}$  with a resolution (unapodized) of 0.25  $cm^{-1}$ ), with hundreds of them sounding the atmospheric temperature. The retrieval becomes an over-constrained problem (more observations than degree of freedom).
- The increased resolution power: with IASI the resolution power is about  $\lambda/d\lambda \simeq 1200$ . Presently, the resolution power of the TOVS (TIROS<sup>1</sup>-N Observational Vertical Sounding) radiometer is between 50 to 100.

It is expected that the vertical resolution and the accuracy of retrievals will substantially increase: the IASI mission requirements are an error of 1 K in atmospheric temperature and 10 % in relative humidity profiles with respectively a 1 Km and 2 Km vertical resolution.

The IASI noise is simulated [Cayla et al., 1995] by a white Gaussian noise (this is a realistic assumption for interferometers) with a Noise Equivalent ( $NE\Delta T$ ) at 280 K (Table 1). The  $NE\Delta T$  at 280 K represents the standard deviation  $st_{280}(\nu)$  of the Gaussian noise for a given wave number  $\nu$ . At a different scene brightness temperature  $T'$ , the standard deviation  $st_{T'}(\nu)$  of the Gaussian noise is computed by:

$$st_{T'}(\nu) = \frac{\frac{\partial B(Tb=280, \nu)}{\partial Tb}}{\frac{\partial B(Tb=T')}{\partial Tb}} \cdot st_{280}(\nu) \quad (2)$$

which shows that the noise level increases as  $T'$  decreases. Figure 1 illustrates the standard-deviation of noise at different  $T'$ . It is expected that these characteristics are an over-estimation of the actual noise level for the instrument. Figure 2 shows the IASI spectrum averaged over the TIGR dataset with the corresponding noise standard deviation spectrum. Note that some spectral regions could have a noise standard deviation larger than 2 K on average.

There are 4 field-of-view for each IASI samples, covering an area of 12 to 9 kms at nadir. Assuming homogeneous meteorological conditions, an average of the 4 pixel measures can be used to perform

<sup>1</sup> Television InfraRed Observation Satellite

the retrievals: these 4 field-of-view provide redundant measurements that can be averaged to reduce noise.

### 3. The Neural Network Inversion Approach

Various neural inversion techniques have been developed, like the "Iterative Inversion" [Kindermann and Linden, 1990], the "Distal Learning" [Jordan and Rumelhart, 1992] or the "Distal Learning" optimized by Monte-Carlo algorithm [Hidalgo and Gómez-Treñño, 1996]. We have chosen to use the "Direct Inversion" approach for two reasons: it performs a global inversion and it is possible to introduce *a priori* information into the method. The *a priori* knowledge is any information about the solution of the problem which is in addition to the information contained in the data set. In usual statistical techniques (like regression), overcoming the "black-box" modeling conception (no assumptions about the physical problem) improves results. Therefore, we have combined three approaches: the *structural stabilization* of the network, regularization of the learning algorithm by the *input perturbation* technique and a physically optimized *feature selection* process in the IASI data. Our numerical experiments have shown that the introduction of this kind of *a priori* information is very useful and makes training possible with relatively few data.

#### 3.1. Global Inversion

In the "Direct Inversion" technique, a MLP neural network is used to estimate directly the mapping between the IASI observations and retrieved geophysical variables. In effect, the "trained" MLP is a statistical model of the inverse RTE, providing once and for all a global inversion. The learning algorithm (the more expensive computational part) is performed off-line only once. Then, the application of the neural network model for the inversion of IASI observations is quasi-immediate in the operational stage: no regressions and no Jacobian computations are required.

Another advantage over classical physico/statistical techniques is that a good initial condition for the inversion is not needed. Moreover the required memory storage is very small. There's also no need for a rapid direct model (necessary in iterative inversion algorithms), where the speed is usually obtained by linearizing the RTE and assuming uncorrelated Gaussian errors.

#### 3.2. MLP and Structural Stabilization of the Architecture

The MLP network is a mapping model composed of parallel processors called "neurons". These processors are organized in distinct layers: the first layer (number 0) represents the input  $X = (x_i; 0 \leq i \leq m_0)$  of the mapping where  $m_0$  is the number of neurons in layer 0. The last layer (number  $L$ ) represents the output of the mapping  $Y = (y_k; 0 \leq k \leq m_L)$ . The intermediate layers ( $0 < m < L$ ) are called the "hidden layers". These layers are connected *via* neuronal links (Figure 3): two neurons  $i$  and  $j$  between two consecutive layers have synaptic connections associated with a synaptic weight  $w_{ij}$ . A neuron executes two simple operations: first, it makes a weighted sum of the inputs and then transfers this signal to its output through a so called transfer, sigmoide of activation function like  $\sigma(a) = \tanh(a)$ . The neuron  $j$  of a hidden-layer of the output-layer has an output  $z_j$

given by:  $z_j = \sigma \left( \sum_{i \in \text{Inputs}(j)} \omega_{ij} \cdot z_i \right)$ . Generally, for

regression problems, the output units have a transfer function that is identity. For example, in a one hidden layer MLP, the  $k^{\text{th}}$  output  $x_k$  of the network is defined as:

$$x_k(y) = \sum_{j \in S_1} w_{jk} \cdot \sigma(a_j) = \sum_{j \in S_1} w_{jk} \cdot \sigma \left( \sum_{i \in S_0} w_{ij} \cdot y_i \right) \quad (3)$$

where  $\sigma$  is the sigmoide function,  $a_j$  is the activity of neuron  $j$  and  $S_i$  is the  $i^{\text{th}}$  layer of the network (with  $i = 0$  for the input layer). We have deliberately omitted the usual bias term in this formula to simplify notation. It has been demonstrated [Hornik et al., 1989; Cybenko et al., 1989] that any continuous function can be represented by a one-hidden layer MLP.

The neuron acts, in its entire input space, as a "fuzzy" linear discriminant: a neuron  $j$  cuts its input space into two half subspaces separated by a plane orthogonal to the vector of its input weights  $\{\omega_{lj}; l \in \text{Inputs}(j)\}$ . On one side of the "frontier" the response of the neuron is 0, on the other side the response is 1 and in the "fuzzy frontier" the response of the neuron is quasi-linear (corresponding to the linear part of the transfer function). So, the MLP network, like linear regression, is very well adapted to high-dimension data because its neurons acts in the entire data space and not in a partition of this space like some methods (radial basis function, splines interpolators, etc).

How is the neural network structure defined? Theoretically, it has been demonstrated [Sontag, 1991] that any inverse problem can be resolved by a two-hidden layer MLP network since such neural networks can take into account discontinuities and extremely nonlinear variations (often present in inverse problems), in contrast to one-hidden layer MLPs that approximate continuous functions.

In practice, the answer can be different: we have observed in our experiments that with noise-corrupted data, a one-hidden layer can be sufficient. Furthermore, our experiments show that smooth solutions shall be obtained by a one hidden layer. This limitation in the number of hidden layers is a structural stabilization: the resulting reduction of the number of free parameters (the synaptic weights  $W$ ) regularizes the neural estimation, producing a functional equivalence between the desired function (the inverse of the RTE) and its estimation (the trained neural network).

### 3.3. Learning Algorithm and Regularization by Input Perturbation

Given an architecture (number of layers, neurons and connections), all the information of the network is contained in the weights  $W$  (the set of all synaptic weights  $\omega_{ij}$ ). The learning algorithm is the optimization technique that estimates the optimal network parameters  $W = \{\omega_{ij}\}$  by minimizing a loss function  $C(W)$  so that the neural mapping approaches as closely as possible the desired function. The most frequently used criterion to adjust  $W$  is the mean square error in network outputs:

$$C(W) = \frac{1}{2} \sum_{k=1}^{m_L} \int \int (y_k(x; W) - t_k)^2 P(t_k/x) P(x) dt_k dx \quad (4)$$

with  $t_k$  the  $k^{th}$  desired output component,  $y_k$  the  $k^{th}$  neural output component and  $P(\cdot)$  the probability density function of input data  $x$ . Practically,  $C(W)$  is approximated by:

$$\bar{C}(W) = \frac{1}{2N} \sum_{e=1}^N (y_k(x; w) - t_k)^2 \quad (5)$$

The Error Back-Propagation (BP) Algorithm [Rumelhart et al., 1986] is used to minimize  $C(W)$ . It is a stochastic steepest descent method very well adapted to this neural architecture because the computational cost is linearly related to the number of parameters.

To reduce the estimation sensitivity to input noise in the data, we use the Input Perturbation technique.

It is an heuristic method to control the effective complexity of the neural network mapping. The technique consists, during the learning step, in adding to each input a random vector representing the instrumental noise. It has been demonstrated [Bishop, 1996] that, under certain conditions (low noise assumption), training with noise is closely related to regularization (or smoothing) technique. In the Input Perturbation method, the usual error function  $C(W)$  (equation (4)) takes the form:

$$\tilde{C}(W) = \frac{1}{2} \sum_{k=1}^{m_L} \int \int \int (y_k(x + \eta; W) - t_k)^2 P(t_k/x) P(x) P(\eta) dt,$$

If the noise  $\eta$  is sufficiently small, we can expand the network function  $y_k(x + \eta; w)$  to first order. Then, we obtain the relation:

$$\tilde{C}(W) \simeq C(W) + \nu \cdot \Omega(W) \quad (7)$$

where  $\nu$  is the noise variance and

$$\Omega(W) = \frac{1}{2} \sum_{i=1}^{m_0} \sum_{k=1}^{m_L} \int \left( \frac{\partial y_k}{\partial x_i} \right)^2 \cdot P(x) \cdot dx \quad (8)$$

is a Tikhonov penalty terms (i.e. stabilizer) which avoids solutions with high gradients (rapid variations of the neural function). So the minimization of this new criterion  $\tilde{C}(W)$  constrains the solutions to be smooth. This regularization technique limits the number of degrees of freedom in the neural network to bring its complexity nearer to the desired function. This limitation reduces the class of possible solutions and makes the solution of the problem unique.

### 3.4. Feature Selection for Dimension Reduction

A MLP neural network can, in principle, be used to map any input vector space to any output vector space; however, in practice, the data representation significantly affects the quality of the final results. In particular, care must be exercised to avoid an over emphasis on the noise component. Dimension reduction techniques can be used to present not only a more compact representation but also more pertinent information to the input of the neural network.

The *curse of dimensionality* stipulates that it is hard to apply a statistical technique to high dimensional data. We have seen in section 3.2 that the MLP is a well-adapted technique in this kind of problem, but practical problems still occur for high dimensional data: for example, the number of parameters (the weights  $W$  in the MLP neural network)

increases with the number of inputs. This can allow excess of degrees of freedom in the neural interpolator which, when combined with the introduction of non-informative data (i.e. noise or spectral information non related to retrieved quantities), may distort the learning process: the quality criterion is more difficult to minimize and the computations are longer.

Thus, the goal of dimension reduction is to present to the neural network the most relevant information from initial rough data (i.e. noisy physical measurements). There exist two ways to reduce the dimension of the input data [Jain and Zongker, 1997]: *feature extraction* (a transformation of rough data by an operator, linear or not) and *feature selection* (selection of channels in input data) [Bishop, 1996]. Feature selection is chosen here: for the retrieval of one geophysical variable, we select channels that are, as far as possible, uniquely sensitive to this one atmospheric parameter. By studying the RTE Jacobians (derivatives of the transmittances with respect to each geophysical parameter), it is possible to analyze mutual information between measured brightness temperatures and geophysical variables [Ch eruy et al., 1993]. But we need to make a compromise between reducing data dimension and preserving the redundant information in rough data to alleviate effects of noise.

## 4. Radiosonde-Based Learning and Test Datasets

### 4.1. Construction of an IASI Learning Data Set: the TIGR Data Base

We use in our application the three TIGR (Thermodynamical Initial Guess Retrieval) data bases of the ARA group: TIGR1 (861 atmospheres) [Ch edin et al., 1985], its 1990 revised version TIGR2 (1761 atmospheres: 322 in tropical air-mass, 388 in mid-latitude type 1, 354 in mid-latitude type 2, 104 in polar type 1 and 593 in polar type 2) [Achard, 1991; Escobar et al., 1993] and its 1997 extended version TIGR3 (2311 atmospheres: same as TIGR2 but with an extended tropical air-mass of 872 atmospheres) [Chevallier et al., 1998]. All of these datasets are constituted from more than 150,000 radiosonde measurements, sampled for their diversity, and described by their temperature and gas concentration profiles with a discretization of atmosphere into 40 layers (see Table 2). This sampling includes a large number of rare events. The final data base is composed of 3494 complex atmospheres. The minimum and maximum envelopes of the TIGR3 atmospheric temperature profiles are

represented in Figure 4 to illustrate the large range of variability that the radiosondes measurement represent. Not only the range of variability can be extreme, but also inversion in the profiles can product complicated structures that are very challenging to any retrieval method.

The 4A (Automatized Atmospheric Absorption Atlas) line-by-line forward radiative transfer algorithm [Scott and Ch edin, 1981; Tournier, 1995] has been used to compute the IASI brightness temperatures associated with these 3494 atmospheres for clear condition over the sea. The 4A algorithm allows for an analytical computation of the physical Jacobians (first derivatives of the transmittance with respect to each variable like temperature, gas concentration, etc) [Ch eruy et al., 1995]. An illustration of such Jacobians versus pressure is given in Figure 5 for the spectral region  $650 - 800 \text{ cm}^{-1}$  ( $15.5 \mu\text{m} - 12.5 \mu\text{m}$ ). The vertical integration of the atmospheric information is illustrated in Figure 6 where Jacobians for 6 wave numbers in the  $15.5 \mu\text{m} - 12.5 \mu\text{m}$  spectral region are shown. Channels with a limited extent (mostly in the lower atmosphere), in terms of vertical resolution, provide a more precise information than the others (in the top of the atmosphere) because a flat Jacobian indicates ambiguities in the retrieved profile. The spacing of the peaks is also important to reduce ambiguities. The concept of vertical resolution depends on both the width and the spacing of the channel's jacobians [Rodgers, 1990].

### 4.2. Improved Representation of the surface temperature in TIGR

In the current TIGR data base, the surface temperature  $T_s$  has been set equal to the temperature of the 40<sup>th</sup> (lowest) atmospheric level T40. This does not represent the actual situation, especially over land, where the surface skin temperature can differ significantly from the near-surface air temperature in systematic ways with time-of-day, latitude, season and location (see for example [Rossow et al., 1989]). For better representativity we statistically generate, for each atmosphere, a set of 10 different  $T_s$  using the T40 information, based on the statistical distribution (i.e. mean and standard deviation) of (T40- $T_s$ ) in a data base of 150,000 radiosonde measurements. Thus, for every atmosphere, knowing T40, we choose randomly 10  $T_s$  with the estimated density probability. For example, in the tropical air-mass, we obtain a  $T_s$  data base of 3220 atmospheres ( $322 \times 10$ ).

## 5. Surface Temperature Retrieval

This study is limited to clear sky oceanic situations and to the tropical air-mass case, emissivity is set equal to 1.0. Ts in the tropical air-mass is very important to climatological analyses.

### 5.1. Jacobian-Based Channel Selection

There are two spectral regions sensitive to the surface characteristics in the IASI spectral domain:  $12.5 \mu\text{m} - 10.2 \mu\text{m}$  ( $\simeq 800 - 980 \text{ cm}^{-1}$ ) and  $4.0 \mu\text{m} - 3.6 \mu\text{m}$  ( $\simeq 2500 - 2750 \text{ cm}^{-1}$ ). It is worth noting that the second spectral region can be contaminated by the sun during the day. However, in these regions, some wavelengths are contaminated by other atmospheric constituents. To eliminate the corrupted channels and to reduce the dimensionality (as explained in section 3.4), we use a channel selection process based upon an analysis of the wavelength sensitivity to Ts variations. We define sensitivity as the mean variation  $I(\nu)$  for 1 K change of Ts (see equation (1)). We select, in these two windows, all channels with a sensitivity higher than a fixed threshold (Figure 7). 357 channels are obtained in the first window (with a threshold of 70 % which realises a good compromise) and 262 in the second window (with a threshold of 85 % because channels are more sensitive to surface temperature in this window).

### 5.2. Network Learning and Testing

The TIGR data base (section 4.2) is divided into a learning base of 3000 atmospheres to make the regression and a base of 220 atmospheres to test the generalization ability of the trained neural mapping.

To retrieve the Ts variable, we use a one-hidden-layer MLP neural network (see section 3.2 for structural stabilization). For the first window ( $800-980 \text{ cm}^{-1}$ ) the neural structure is 357-20-1: 357 neurons in the input layer (357 selected brightness temperatures), 20 neurons in the hidden layer and 1 neuron in the output layer (representing Ts). For the second window ( $2500-2750 \text{ cm}^{-1}$ ), the structure is 262-20-1.

This neural mapping is then trained by the Error Back-Propagation algorithm on the learning base. The Input Perturbation regularization technique is used: simulated noise (according to the NE $\Delta$ T specifications) is added to the input data during the learning step. The generalization ability of our model was then tested on noisy data computed on the 220 test atmospheres. The instantaneous retrieval of Ts from noisy data gives a generalization RMS of approxi-

mately 0.4 K. Similar results are obtained using only the second spectral window. Without noise, the RMS error is less than 0.3 K, i.e., the retrieval error is significantly affected by measurement error, not the error of the neural regression fit.

## 6. Atmospheric Temperature Profile Retrieval

The 40 layers of 4A (see Table 2) were used to compute the brightness temperature spectrum for clear condition over ocean, but for the retrieval, the vertical discretization of the atmosphere has been changed (from 4A-levels to 1Km-levels) to match IASI specifications. The objective of this section is then to retrieve the 32 lower atmospheric temperature of the 1Km-layer profiles.

### 6.1. Channel Selection

The choice of the channels for the retrieval of temperature profiles is made so that they are, as much as possible, sensitive to only one constant-concentration gas; then, variations of  $I(\nu)$  in equation (1) result mainly from temperature variations. Thus, the “ $\text{CO}_2$ ,  $\text{NO}_2$  (or both) absorbing-spectral regions” are used for the retrieval of atmospheric temperature profiles: the  $15.5 \mu\text{m} - 12.5 \mu\text{m}$  ( $\simeq 645 - 800 \text{ cm}^{-1}$ ) and the  $4.7 \mu\text{m} - 4.0 \mu\text{m}$  ( $\simeq 2100 - 2500 \text{ cm}^{-1}$ ) spectral regions.

To present the most relevant information to the neural network inputs (section 3.4), we use a channel selection process. The feature selection method is based on the study of the Jacobians in order to define the sensitivity of a channel to atmospheric temperature. The mean Jacobian in TIGR3 indicates the sensitivity relation between atmospheric layers and channels. The standard deviation of the Jacobian (around the mean) is negligible except near the surface; this means that the mean Jacobian is robust to the atmospheric situation except in the lower atmospheric layers.

The feature selection process has two steps. First, channels are selected which satisfy quality criteria, i.e. specifying, as unambiguously as possible: (1) the Jacobian extent of a channel (characterized by the area below the Jacobian) and the Jacobian width at mid-height have to be smaller than fixed thresholds; (2) the Jacobian center of a channel is not near surface; and (3) the Jacobian has a single peak. For the  $15.5 \mu\text{m} - 12.5 \mu\text{m}$  spectral region, we have selected 442 channels within the 621 channels of the spectral

range ( $645 - 800 \text{ cm}^{-1}$  with  $0.25 \text{ cm}^{-1}$  resolution).

The second step chooses a vertically uniform subset of the channels that meet the quality criteria. The IASI instrument gives little information below 10 hPa, so our retrievals will be limited to the pressure range 1013-10 hPa (32 layers with discretization of 1 Km). We have chosen 9 channels for each of 30 layers (the previous 32 layers minus the two lowest layers sensitive to surface temperature) between 1013 and 10 hPa. The final number of channels is 270. However, it is important to note that the layers 23-28 have a deficit in channels and that the sensitivity is higher in the lower atmosphere (Figure 8).

The  $4.7 \mu\text{m} - 4.0 \mu\text{m}$  spectral region is also important for the atmospheric temperature profile retrieval for two reasons. First, the lower atmospheric Jacobians are narrower than in the  $15.5 \mu\text{m} - 12.5 \mu\text{m}$  region allowing for a better vertical resolution. Second, the channels are less affected by water vapor.

However, due to the larger noise in this spectral domain, the channel selection has to be performed differently than in the  $15.5 \mu\text{m} - 12.5 \mu\text{m}$  region. The IASI noise (see section 2.2) - the standard deviation of the Gaussian noise - may be as large as a few degrees for channels sensing the higher layers (low brightness temperatures). The redundancy of the information due to the number of channels doesn't compensate this noise. Consequently, the spectral range used covers mainly the lower atmospheric layers. The Jacobian analysis selects channels in the  $2140 - 2240 \text{ cm}^{-1}$  spectral range (401 channels).

## 6.2. Network Learning and Testing

All the atmospheres of the learning and the test bases are described by 30 atmospheric temperatures (4A levels up to 7 hPa for 32 Km height) and the corresponding 671 selected brightness temperatures computed by 4A. The neural network structure used for the regression is then 671-50-30: 671 units in the input layer (the 671 selected channels in the  $15.5 \mu\text{m} - 12.5 \mu\text{m}$  and the  $4.7 \mu\text{m} - 4.0 \mu\text{m}$  spectral regions), 50 units in the hidden layer and 30 units in the output layer (the 30 lower atmospheric temperatures in 4A-levels, the interpolation to 32 1Km-level being made a posteriori).

We have tested four configurations: for the "All-air-masses" and the "Tropical-air-mass", with and without the 4-pixels averaging (noise divided by 2, see section 2.2).

**6.2.1. "All-air-masses" configuration.** We have merged the TIGR1 and TIGR3 data bases of section 4.1 and the resulting 3155 atmospheres have been randomly subdivided into a learning base of 2700 atmospheres and a test base of 455 atmospheres.

The RMS fit errors (given for the 32 atmospheric 1Km-layers) for the learning and the test sets are shown in Figure 9-A for the 1-pixel configuration and Figure 9-B for the 4-pixels configuration. We have an overall good agreement between the computed and "observed" temperature profiles: rms errors close to 1 K on average (less than lower than 1.3 K except near 10 hPa). Also, we can see that we are facing some problems in two vertical regions:

- In the upper layers of the atmosphere: IASI provides poor information above 20 hPa (see Figure 5) due to the fact that the Jacobians of the channels sounding these layers are more vertically extensive than channels near surface and their amplitudes are smaller. So, the compensation phenomenon is more important in this vertical region. Some of our experiments have shown that the addition of the AMSU/A (also planned for flight on board METOP-1) information improves results in this vertical region.
- In the near-surface layers: the difference  $T_{40} \neq T_S$  complicate the retrieval due in part to the compensation phenomenon (a under-estimation of temperature in one layer is compensated by an over-estimation in a near-by layer). Considerations about specific neural networks compensation phenomenon are given in [Aires et al., 1999; Aires, 1999]. It is possible that the simultaneous retrieval of  $T_S$  and  $T_{40}$ , being more constrained, may solve this problem.

Thus, even though the TIGR database possesses atmospheric situations with highly variable temperature profiles, the RMS errors obtained in Figures 9-A and 9-B are close to the IASI objective (1 K of RMS error for 1 Km in vertical resolution).

The use of 4-pixel averages uniformly decreases (by about 0.1 K) the RMS in the atmospheric layers. This relatively small improvement is due to the fact that the solution regularization used to avoid noise effects by the input perturbation method is sufficiently efficient, so the reduction of noise by pixel-averaging has a reduced impact on the quality of the retrievals. This fact means that our method is able to provide good results for each pixel to maximize the horizon-

tal resolution or to perform scene selection. Five randomly chosen examples of retrievals in the test set are shown in Figure 10.

**6.2.2. "Tropical-air-mass" configuration.** We have merged the Tropical-TIGR1 and the Tropical-TIGR3 data bases of section 4.1 and the resulting 1070 atmospheres have been randomly subdivided into a learning base of 1000 atmospheres and a test base of 70 atmospheres.

The RMS errors (given for the 32 atmospheric 1Km-layers) in the learning and the test set are given in the Figure 9-C for the 1-pixel configuration and in the Figure 9-D for the 4-pixels configuration. We see that the RMS profile is significantly improved at 1 K, so the specialization of the neural network to the tropical air-mass is important. As above, the RMS is also decreased by about 0.1 K with the 4-pixels average configuration.

It is important to note that the specialization of the neural network on one air mass:

- improves the retrievals;
- requires a training data base with a larger number of atmospheres.

In this case, the 1070 tropical atmospheres are not sufficient, so differences between the learning and the test bases are not negligible. Future work should address this very important problem of the full representativity of the learning and testing bases.

## 7. Conclusion and Perspectives

A neural network approach uses a maximum of *a priori* information to limit the number of free parameters in the neural model so as to constrain the retrieval of surface and atmospheric temperatures as a "better-posed" problem. The method is trained using the TIGR data base, i.e. a vast and complex set of atmospheric situations (from radiosonde measurements which are much more irregular than model output) with a wide range of radiosonde conditions including rare events. This fact is important to judge the quality of our results. The surface temperature for tropical situations displays a RMS error of 0.4 K for instantaneous retrievals). Results for atmospheric temperature profile retrievals are given for four configurations ("All-air-masses" or "Tropical-air-mass", with and without the 4-pixels average). Results are close to the specifications of the WMO for the "All-air-masses" configurations: 1 K of error for the instan-

aneous temperature retrieval with 1 Km vertical resolution. The specialization to the "Tropical-air-mass" significantly improves the results, which means that using a specialized neural network for a few different air-masses is the good strategy, but a larger dataset is then required to train these specialized models. It is important to note that the results obtained for the IASI retrievals and entirely depend on the complexity of the dataset used to perform the statistics. Thus, it has been demonstrated in this work the potential of the IASI instrument to achieve the WMO specifications for realistic conditions even for the complex situations included here. This new instrument is a clear advance over current instruments. The MLP inversion technique developed here for the processing of IASI observations is flexible enough to introduce a priori information in the retrieval scheme, is robust to noise, and is accurate and very fast.

We plan to use independently a neural network for the two other air-masses (temperate and polar) by increasing the TIGR data base. Another idea is to use this methodology with more channels so as to retrieve not only the surface temperature and the temperature profile, but also water vapor and ozone profiles. The simultaneous retrieval of these variables is expected to exploit the correlations between variables in order to better constrain the inversion process. Considerable improvements are expected by the use in parallel of AMSU/A observations. Finally, further improvement may also be expected by the introduction of a first-guess solution in the MLP inversion [cf. Aires *et al.*, 2000].

**Acknowledgments.** We would like to thank the CNES and Thomson-CSF for supporting parts of this research. All computations were made in a CRAY C90 of the IDRIS (Institut du Développement et des Ressources en Informatique Scientifique) computational center of the C.N.R.S. (Centre National de la Recherche Scientifique).

## References

- Achard, V., Trois problèmes clés de l'analyse tridimensionnelle de la structure thermodynamique de l'atmosphère par satellite : mesure du contenu en ozone, classification des masses d'air, modélisation hyper-rapide du transfert radiatif, PhD thesis, Université Pierre et Marie Curie (Paris VI), 1991.
- Aires, F., R. Armante, A. Chédin, and N.A. Scott, Surface and atmospheric temperature retrieval with the high resolution interferometer IASI, *Proc. of the Am. Meteorol. Soc.*, 98,, Paris, No P5.46.B, 181-186, 1998.
- Aires, F., Problèmes inverses et réseaux de neurones: Ap-

- plication à l'interféromètre haute résolution IASI et à l'analyse de séries temporelles, *PhD Thesis*, Université Paris IX/Dauphine, CNES Report, 1999.
- Aires, F., M. Schmitt, N.A. Scott, and A. Chédin, The Weight Smoothing regularization for Jacobian stabilization, *IEEE Trans. on Neural Networks*, **10**, 6, 1502-1510, 1999.
- Aires, F., C. Prigent, W.B. Rossow, and M. Rothstein, A new neural network approach including first-guess for retrieval of atmospheric water vapor, cloud liquid water path, surface temperature and emissivities over land from satellite microwave observations, *Journal of geophysical research*, in press, 2000.
- Amato, U., and C. Serio, The impact of ransom noise on the performance of a new infrared atmospheric sounding interferometer (IASI) configuration, *Int. J. Remote Sensing*, **18**, 15, 3,135-3,143, 1997.
- Badeva, V. and V. Morozov, Problèmes incorrectement posés, Masson, 1991.
- Bishop, C., *Neural networks for pattern recognition*, Clarendon Press, 1996.
- Cayla, F., B. Tournier, and P. Hebert, Performance budgets of IASI options, Technical Report IA-TI-0B-5476-CNE, CNES, 1995.
- Chédin, A., N.A. Scott, C. Wahiche, and P. Moulinier, The improved initialization inversion method: a high resolution physical method for temperature retrievals from TIROS-N series, *J. Clim. Appl. Meteor.*, **24**, 128-143, 1985.
- Chéruy, F., N.A. Scott, R. Armante, B. Tournier, and A. Chédin, Contribution to the development of radiative transfer models for high spectral resolution observations in the infrared, *J. Quant. Spectrosc. Radiat. Transfer*, **53**, 6, 597-611, 1995.
- Chéruy, F., J. Susskind, N.A. Scott, A. Chédin, and J. Joiner, GLA and LMD approaches to the processing of AIRS and IASI observations, In *NATO ASI Series: High Spectral Resolution Infrared Remote Sensing for Earth's Weather and Climate Studies*, Springer-Verlag, Vol. I 9, 101-112, 1993.
- Chevallier, F., J.-J. Morcrette, F. Chéruy, and N.A. Scott, Use of a neural network-based longwave radiative transfer scheme in the ECMWF atmospheric model, *Q. J. R. Meteor. Soc.*, **126**, 761-776, 2000.
- Chevallier, F., F. Chéruy, N.A. Scott, and A. Chédin, Neural network approach for a fast and accurate computation of the LongWave radiation budget, *J. Applied Met.*, **37**, 1385-1397, 1998.
- Cybenco, G., Approximation by Superpositions of a Sigmoidal Function, *Math. Control Signals Systems*, **2**, 303-314, 1989.
- Escobar, J., A. Chédin, F. Chéruy, and N.A. Scott, Réseaux de neurones multicouches pour la restitution de variables thermodynamiques atmosphériques à l'aide de sondes verticales satellitaires, *C. R. Acad. Sci. Paris*, **317**, 2, 911-918, 1993a.
- Escobar, J., Base de données pour la restitution de paramètres atmosphériques à l'échelle globale ; étude sur l'inversion par réseaux de neurones des données des sondes verticales atmosphériques satellitaires présents et à venir, *PhD Thesis*, Université Denis Diderot (Paris VII), 1993b.
- Hidalgo, H., and E. Gómez-Treviño, Application of constructive learning algorithms to the inverse problem, *IEEE Transactions on geoscience and remote sensing*, **34**, 4, 1996.
- Hornik, K., M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, *Neural Networks*, **2**, 359-366, 1989.
- Jain, A., and D. Zongker, Feature selection: evaluation, application, and small sample performance, *IEEE transactions on pattern analysis and machine intelligence*, **19**, 2, 153-158, 1997.
- Jordan, M., and D. Rumelhart, Forward models : supervised learning with a distal teacher, *Cognitive Science*, **16**, 307-354, 1992.
- Kindermann, J., and A. Linden, Inversion of neural networks by gradient descent, *Parallel Computing*, 270-286, 1990.
- Prunet, P., J.-N. Thépaut, and V. Cassé, The information content of clear sky IASI radiances and their potential for numerical weather prediction, *Q. J. R. Meteorol. Soc.*, **124**, 211-241, 1998.
- Rodgers, C.D., Characterization and error analysis of profiles retrieved from remote sounding measurements, *Journal of Geophysical Research*, **95**, D5, 5587-5595, 1990.
- Rodgers, C.D., Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation, *Review of Geophysics and Space Physics*, **14**, 609-624, 1976.
- Rossow, W.B., C.L. Brest, and L.C. Garder, Global, seasonal surface variations from satellite radiance measurements, *J. of Climate*, **2**, 214-247, 1989.
- Rumelhart, D., G. Hinton, and R. Williams, Learning representations by back-propagating error, volume I, chapter 8, 318-362, 1986.
- Scotts, N.A., A. Chédin, A fast line-by-line method for atmospheric absorption computations: the automatized atmospheric absorption atlas, *J. of Applied Meteorology*, **20**, 20 802-20812, 1981.
- Sontag, E., Feedback stabilisation using two-hidden-layer nets, Technical Report 11, Rutgers Center for Systems and Control, 1991.
- Tournier, B., R. Armante, and N.A. Scott, Stransac-93, 4A-93: Développement et validation des nouvelles versions des codes de transfert radiatif pour application au projet IASI, Technical Report 201, ARA/LMD Group, 1994.
- Twomey, S., Introduction to the mathematics of inversion in remote sensing and indirect measurements, Elsevier scientific publishing company, 1977.

---

F. Aires, Department of Applied Physics, Columbia University, NASA Goddard Institute for Space Studies, 2880 Broadway, New-York, NY 10025, USA. (e-mail: faires@giss.nasa.gov)

W.B. Rossow, NASA Goddard Institute for Space Studies, 2880 Broadway, New-York, NY 10025, USA. (e-mail: wrossow@giss.nasa.gov)

A. Chédin and N.A. Scott, Laboratoire de Météorologie Dynamique, École Polytechnique, 91128 Palaiseau Cedex, France. (e-mail: chedin@jungle.polytechnique.fr)

... .., 2000; revised ....., 2000. accepted ....., 2000.

---

<sup>1</sup>Now at Department of Applied Physics, Columbia University, NASA Goddard Institute for Space Studies, New-York

Table 1.

$\nu$ in $cm^{-1}$	$NE\Delta T$ in K	$\nu$ in $cm^{-1}$	$NE\Delta T$ in K
650	0.28	1660	0.34
770	0.28	2090	0.5
790	0.34	2100	0.5
980	0.34	2420	0.5
1000	0.28	2430	0.6
1070	0.28	2500	0.77
1080	0.34	2600	1.1
1200	0.34	2700	1.58
1210	0.28	2760	1.97
1650	0.28		

**Table 2.** The 40 pressure levels of the 4A algorithm

Level	Pressure in hPa	Altitude in Km	Level	Pressure in hPa	Altitude in Km
1	0.05	68.4	21	131.20	14.1
2	0.09	64.3	22	161.99	12.6
3	0.17	59.9	23	200.00	11.1
4	0.30	56.0	24	222.65	10.4
5	0.55	51.8	25	247.90	9.7
6	1.00	47.7	26	275.95	8.9
7	1.50	44.9	27	307.20	8.2
8	2.23	42.2	28	341.99	7.4
9	3.33	39.4	29	380.73	6.7
10	4.98	36.6	30	423.85	6.0
11	7.43	33.9	31	471.86	5.2
12	11.11	31.1	32	525.00	4.5
13	16.60	28.3	33	584.80	3.7
14	24.79	25.6	34	651.04	3.0
15	37.04	22.8	35	724.73	2.3
16	45.73	21.3	36	800.00	1.6
17	56.46	19.9	37	848.69	1.2
18	69.71	18.4	38	900.33	0.8
19	86.07	17.0	39	955.12	0.4
20	106.27	15.5	40	1013.00	0.0

- Figure 1.** Standard Deviation of IASI instrument noise for different brightness temperature measurement  $T'$
- Figure 2.** Mean IASI spectrum (left) and corresponding standard deviation of IASI instrumental noise (right)
- Figure 3.** Architecture of a MLP neural network with  $L$  layers, with inputs  $X$  and outputs  $Y$
- Figure 4.** Minimal and maximal envelope of TIGR3 atmospheric temperature profiles for A "All-Air-Masses", B "Tropical-Air-Mass", C "Temperate 1 Air-Mass", D "Temperate 2 Air-Mass", E "Polar 1 Air-Mass", and F "Polar 2 Air-Mass"
- Figure 5.** Mean (for TIGR3 atmospheres) temperature Jacobian in the  $15.5 \mu m - 12.5 \mu m$  spectral region
- Figure 6.** Atmospheric temperature Jacobian profile for IASI and for 6 channels in the  $15.5 \mu m - 12.5 \mu m$  spectral region
- Figure 7.** Sensitivity in % to a 1 K-perturbation of surface temperature versus wave number, in the two IASI windows ( $800-980 \text{ cm}^{-1}$  and  $2500-2750 \text{ cm}^{-1}$ )
- Figure 8.** Mean temperature Jacobian on TIGR3 for the 270 channels (ordered by maximum absorption altitude) selected on the  $15.5 \mu m - 12.5 \mu m$  spectral region
- Figure 9.** RMS error profile for the atmospheric temperature retrieval in the learning set (continuous line) and in the generalization set (discontinuous line) : A for configuration "All-air-masses/1 pixel", B for configuration "All-air-masses/4 pixels", C for configuration "Tropical-air-mass/1 pixel", and D for configuration "Tropical-air-mass/4 pixels".
- Figure 10.** Five atmospheric temperature profile retrieval examples in the configuration "All-air-masses/1 pixel"

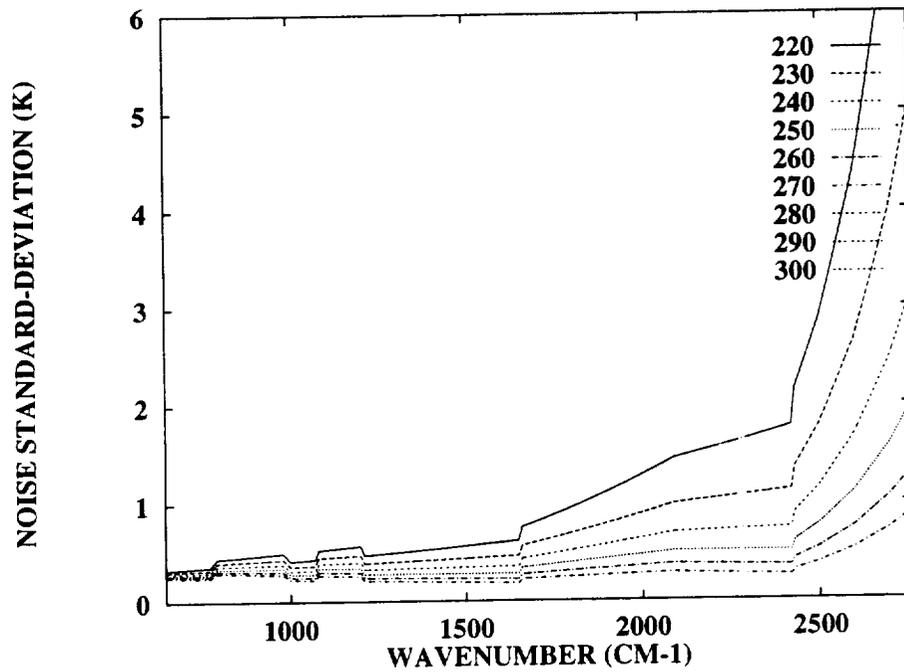


Figure 1.

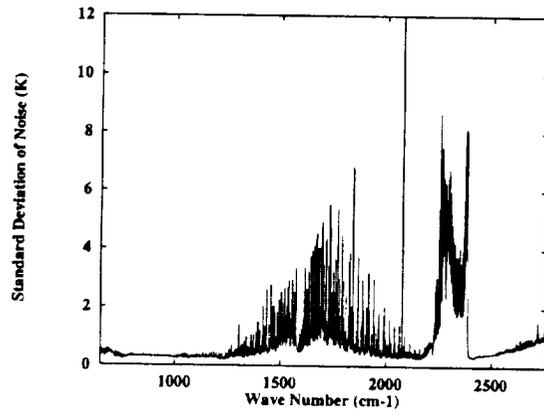
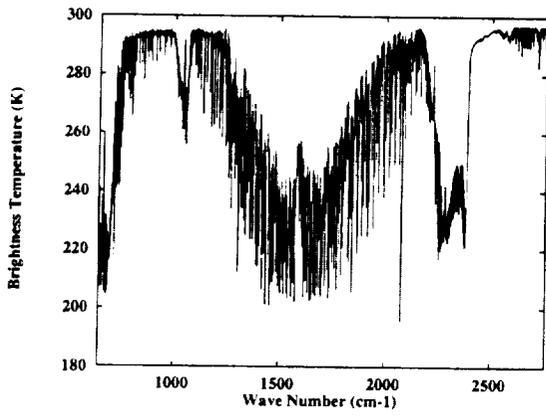


Figure 2.

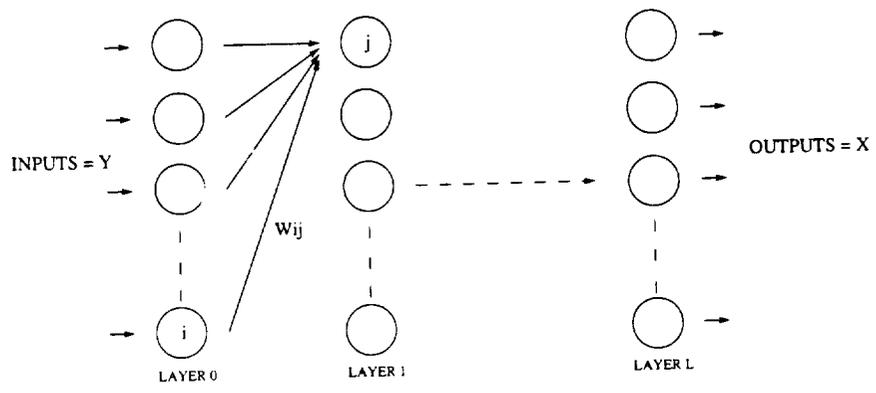


Figure 3.

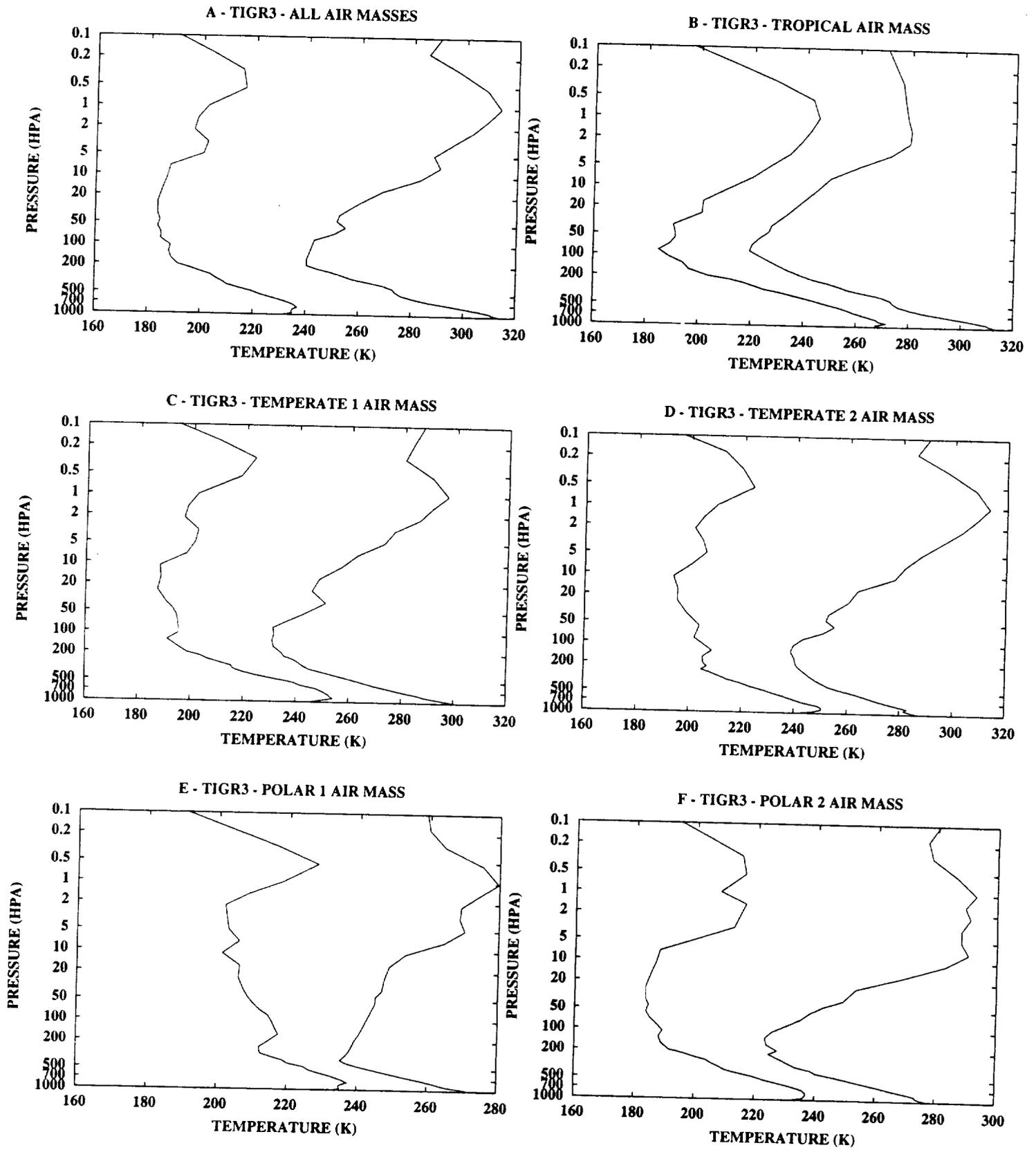


Figure 4.

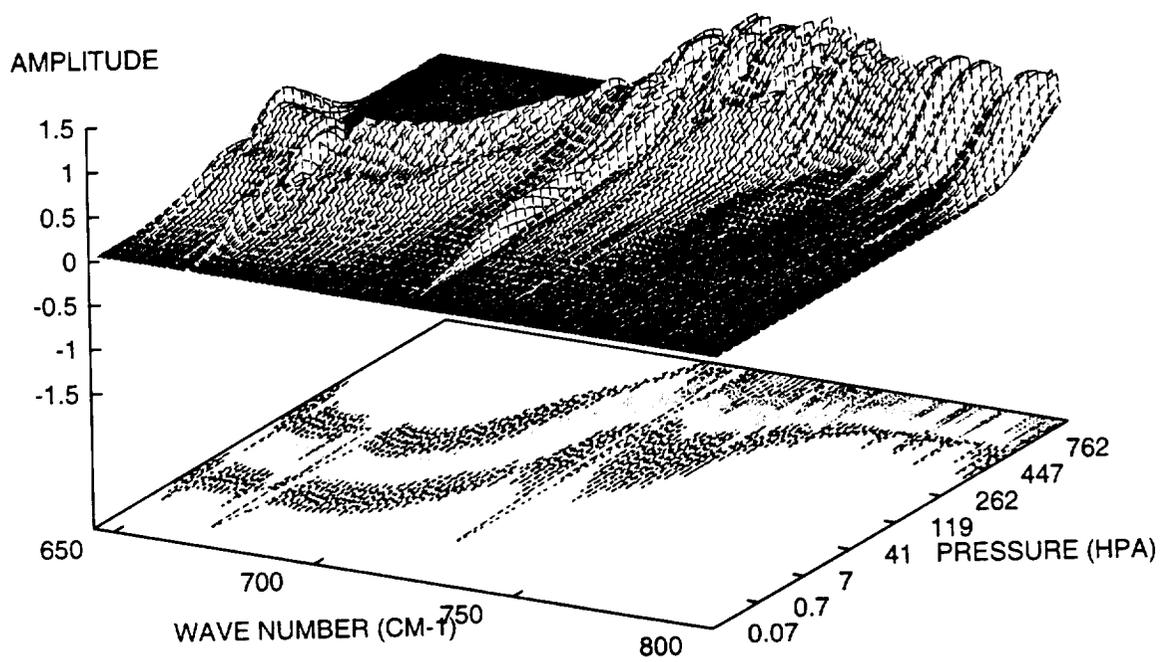


Figure 5.

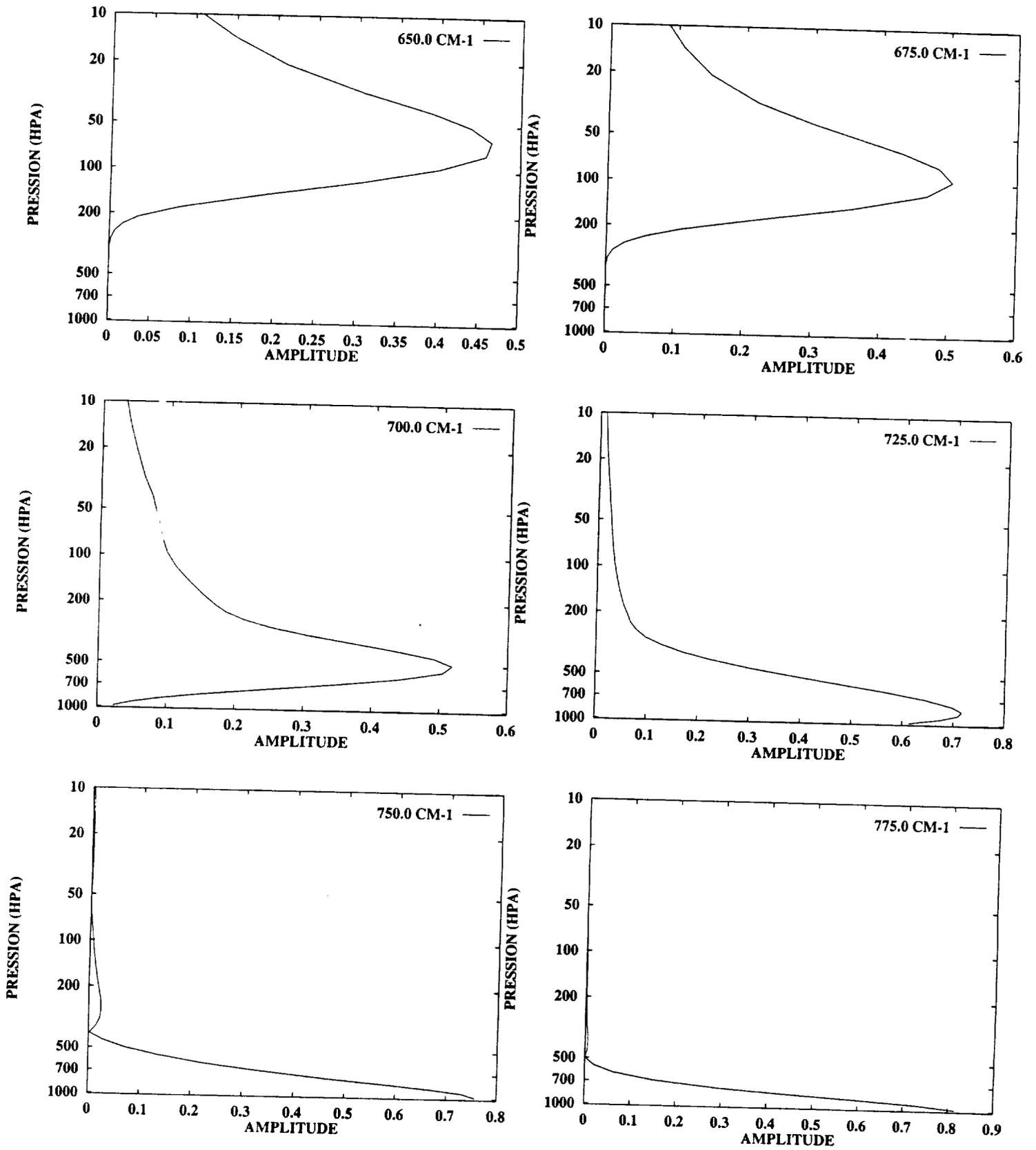


Figure 6.

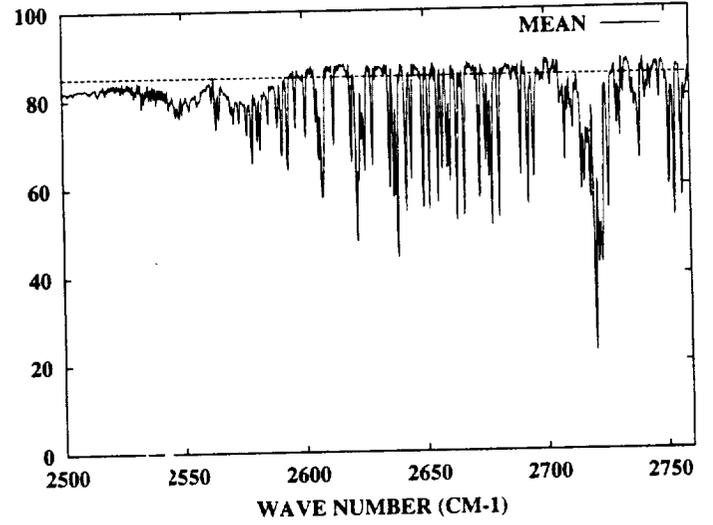
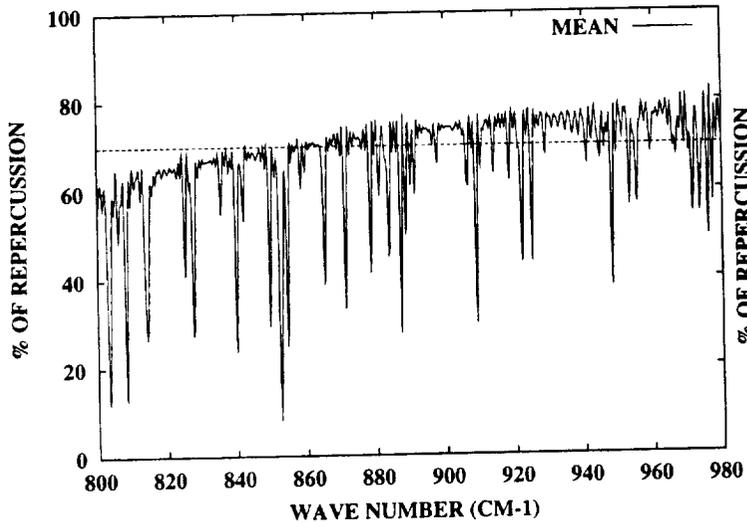


Figure 7.

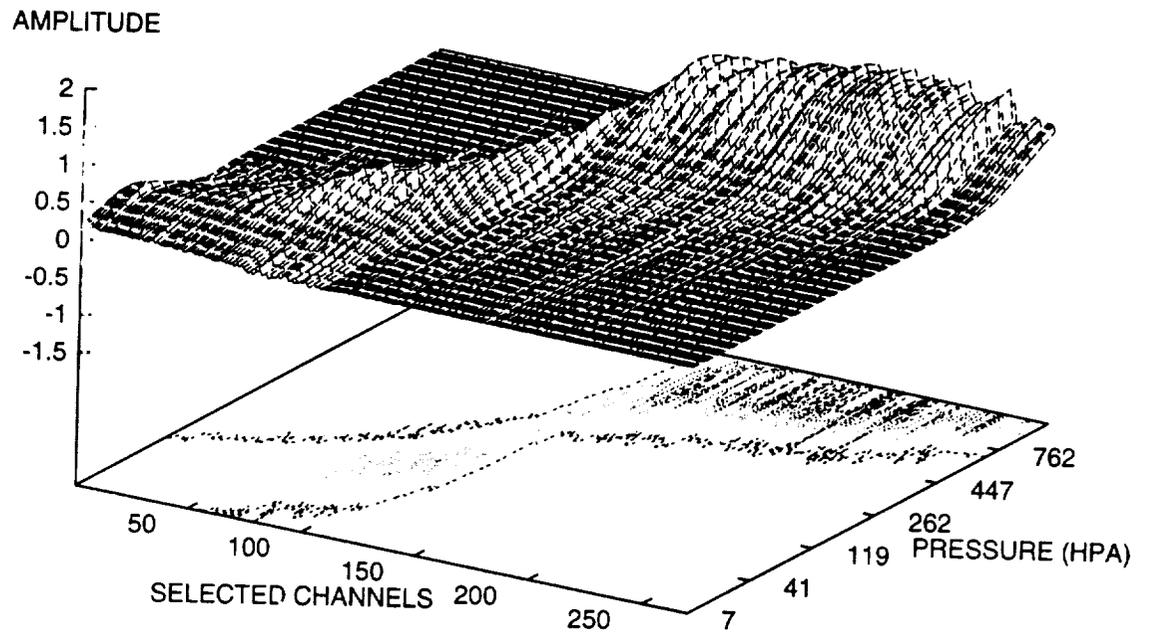


Figure 8.

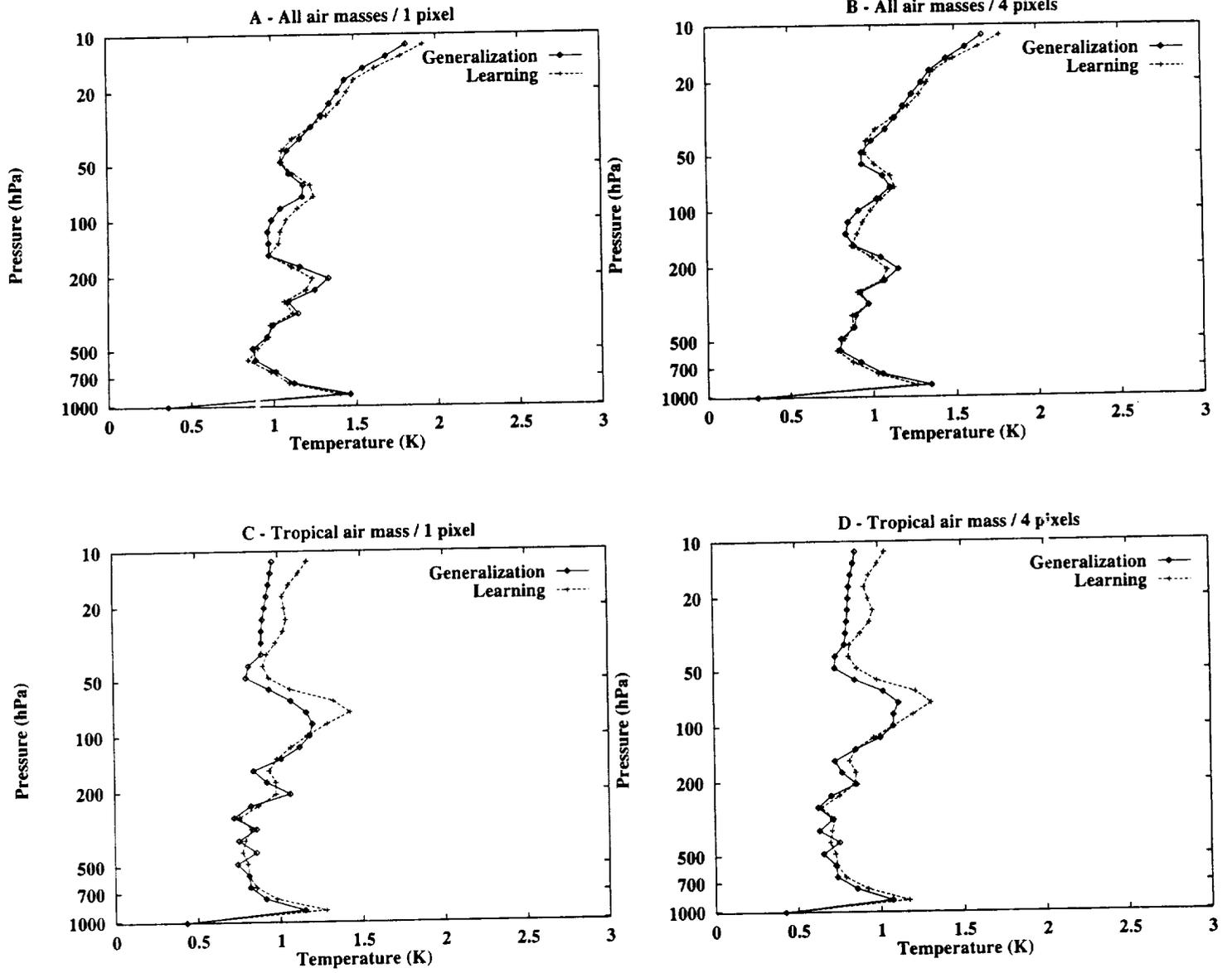


Figure 9.

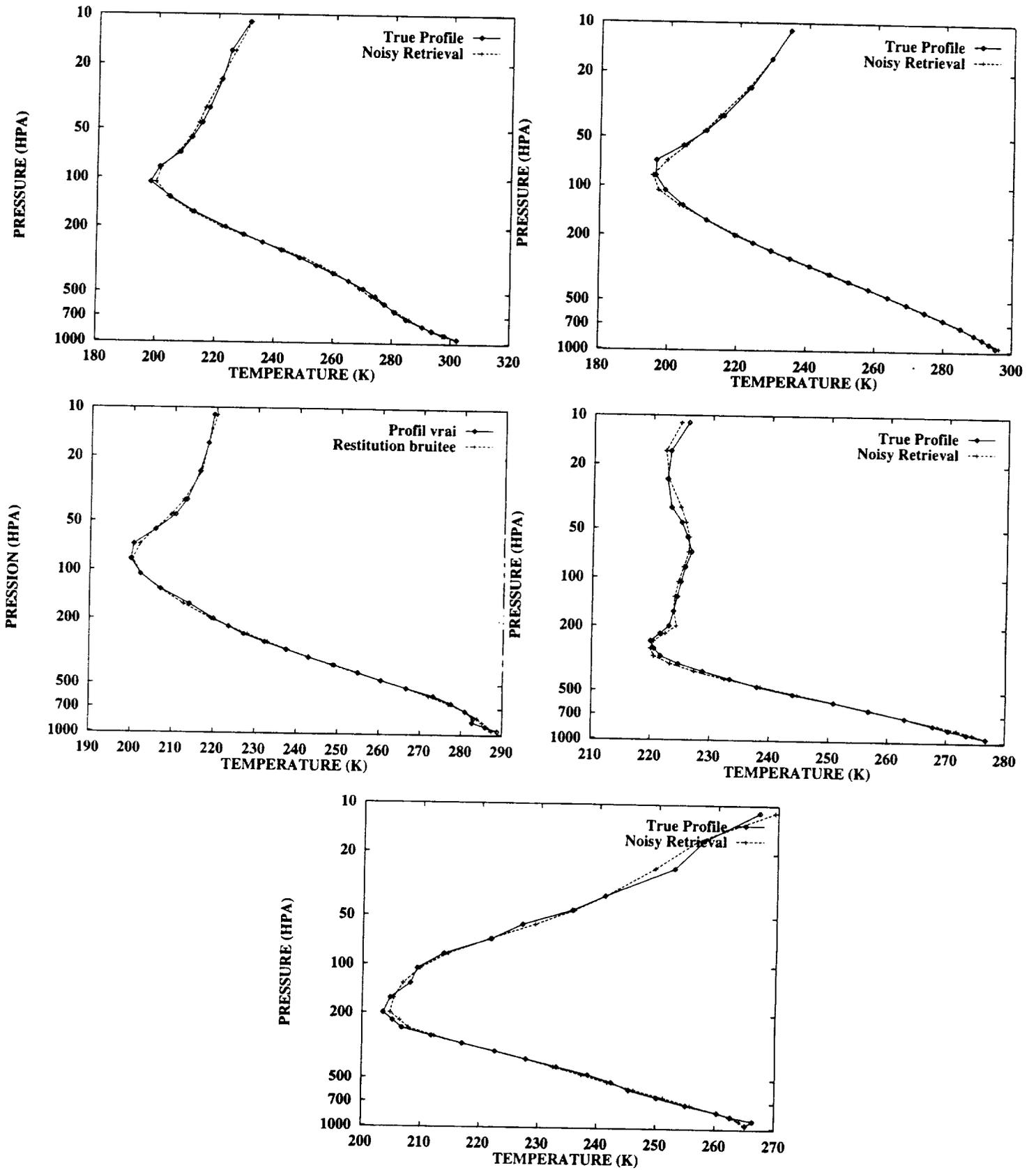


Figure 10.